Statistika

(MD360P03Z, MD360P03U)
ak. rok 2007/2008

Karel Zvára

karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara

(naposledy upraveno 12. listopadu 2007)

proč **testování hypotéz**

- připomeňme 95% intervaly spolehlivosti pro šestku u kostek:
  - kostka A: (0,10; 0,24)
  - kostka B: (0,31; 0,51)
- znamená něco, když $1/6 = 0{,}167$ leží či neleží v 95% intervalu spolehlivosti?
- nelze bezpečně poznat, že kostka A není falešná nebo že kostka B je falešná
- intervaly spolehlivosti určily rozmezí, kde by skutečná pravděpodobnost šestky měla být, jejich spolehlivost je velká, ale omezená
- musíme připustit, že jsme mohli mít smůlu, že se v našich pokusech náhodou realizovaly málo pravděpodobné možnosti, přestože k takové smůle dochází jen zřídka
- potřebujeme **standardizovaná pravidla**, jak rozhodovat

hypotézy a možná rozhodnutí

- možné statistické **hypotézy**
  - **(nulová) hypotéza** H_0: – zjednodušuje situaci, zpravidla se jí snažíme vyvrátit, abychom věcně něco prokázali: porovnávané populace se **ne**liší, vyšetřované znaky jsou **ne**závislé . . .
    tedy žádný (tj. **nulový**) rozdíl, žádná (tj. **nulová**) závislost
  - **alternativa** H_1: **(alternativní hypotéza)** – opak nulové hypotézy, zpravidla to, co chceme věcně dokázat
- možná **rozhodnutí**
  - **zamítnout** H_0 pokud naše data svědčí proti H_0
  - **nezamítnout** H_0 (přijmout H_0) pokud *není dost důvodů* H_0 zamítnout
- hypotéza – tvrzení o **populaci**
- rozhodujeme na základě dat z **výběru**
- nelze zaručit bezchybnost rozhodnutí

chyby v rozhodování

- nelze zaručit bezchybnost rozhodnutí, mohou nastat chyby:
  - **chyba 1. druhu**, když zamítneme platnou hypotézu H_0
  - **chyba 2. druhu**, když nepoznáme, že hypotéza H_0 neplatí a nezamítneme ji (přijmeme ji)
- nechceme příliš často *chybně* zamítat H_0 (tedy falešně něco věcně prokazovat)
- proto se snažíme chybě 1. druhu pokud možno vyvarovat, nelze ji vyloučit
- **hladina testu** α = maximální přípustná pravděpodobnost chyby 1. druhu (zpravidla $\alpha = 0{,}05$, tj. $\alpha = 5$ %)
- **síla testu** = pravděpodobnost správného zamítnutí neplatné hypotézy

schéma rozhodování

rozhodnutí	H_0 platí	H_0 neplatí
H_0 zamítnout	chyba 1. druhu (pst $\leq \alpha$) hladina testu	správné rozhodnutí (pst $1-\beta$) síla testu
H_0 nezamítnout (přijmout)	správné rozhodnutí (pst $\geq 1-\alpha$)	chyba 2. druhu (pst β)

- volíme řádek
- nevíme, který sloupec platí

klasický postup při rozhodování

- zvolit (nulovou) hypotézu H_0, alternativu H_1
- zvolit hladinu testu α
- zvolit metodu rozhodování (který test použít)
- z dat spočítat testovou statistiku T a porovnat ji s tabelovanou kritickou hodnotou (bude ještě: porovnat p-hodnotu s hladinou α)
- **kritický obor** – množina těch výsledků pokusu (např. hodnot T), kdy budeme hypotézu zamítat
- když padne statistika T do **kritického oboru**, pak hypotézu zamítnout (zpravidla, když $T \geq t_0$, t_0 – kritická hodnota)

příklad: padá na kostce šestka příliš často?

- chceme na 5% hladině prokázat, že pravděpodobnost šestky na dané kostce je větší, než by měla být (tj. větší než 1/6)
- $H_0 : \mathrm{P}(\text{padne šestka}) = 1/6 \qquad (\pi = \pi_0)$
- $H_1 : \mathrm{P}(\text{padne šestka}) > 1/6 \qquad (\pi > \pi_0)$
- provedeme $n = 100$ pokusů, Y je počet šestek
- co svědčí pro neplatnost hypotézy? Je to situace, kdy „šestka padá mnohem častěji, než by měla padat za H_0“
- **tvar kritického oboru**: hypotézu zamítat, když $Y \geq y_0$
- za platnosti H_0 má počet šestek Y rozdělení $\mathrm{bi}(n, 1/6)$
- **velikost kritického oboru**: y_0 zvolíme tak, abychom hypotézu za její platnosti zamítali s pravděpodobností nejvýše α, tj.

$$\mathrm{P}_0(Y \geq y_0) \leq \alpha$$

příklad: jak zvolit kritickou hodotu y_0?

- některé pravděpodobnosti pro $Y \sim \mathrm{bi}(100, 1/6)$

y_0	20	21	22	23	**24**	25
$\mathrm{P}(Y \geq y_0)$	0,220	0,152	0,100	0,063	**0,038**	0,022

- podmínku $\mathrm{P}(Y \geq y_0) \leq 0{,}05$ splňuje $y_0 = 24$
- padne-li ve 100 nezávislých hodech kostkou aspoň 24 šestek, budeme na **5% hladině zamítat hypotézu**, že pst šestky je 1/6 **ve prospěch alternativy**, že pst šestky je větší než 1/6 (dáno zvolenou alternativou)
- na kostce A nám padlo 17 šestek, hypotézu **nezamítáme**, to ale neznamená, že bychom hypotézu prokázali
- na kostce B nám padlo 41 šestek, hypotézu **zamítáme**
- pro $\alpha = 10$ % bychom zvolili $y_0 = 22$, bylo by však větší riziko zamítnutí platné hypotézy

příklad: síla testu

- **síla testu** = pst, že hypotézu zamítneme, když ona neplatí
- při 100 hodech hypotézu na 5% hladině zamítáme, je-li $Y \geq 24$
- nechť je ve skutečnosti $\pi = 1/4$, pak hypotézu zamítneme (výsledek pokusu padne do kritického oboru) s pstí

$$\mathrm{P}(Y \geq 24) = \sum_{k=24}^{100} \binom{100}{k} \left(\frac{1}{4}\right)^k \left(1 - \frac{1}{4}\right)^{100-k} = 0{,}629$$

- pro $\pi = 0{,}25$ je tedy síla testu 62,9 %
- pro $\pi = 0{,}3$ je podobně síla testu rovna 92,4 %
- pro $\pi = 0{,}2$ je podobně síla testu rovna 18,9 %

rozhodování pomocí *p*-hodnoty

- ***p*-hodnota** p je nejmenší α, při kterém H_0 z daných dat ještě zamítáme
- p-hodnota p je za platnosti H_0 spočítaná *pravděpodobnost* výsledků stejně nebo *méně příznivých* pro H_0
- H_0 zamítáme právě tehdy, když je $\boxed{p \leq \alpha}$
- p-hodnotu počítají moderní počítačové programy
- existují úlohy, kdy se rozhoduje pouze podle p-hodnoty (např. Fisherův exaktní test ve čtyřpolní tabulce)
- statistické rozhodování:
  spočítat k T odpovídající p-hodnotu a porovnat ji s α

příklad: rozhodování pomocí *p*-hodnoty

- snažíme se prokázat, že šestka padá příliš často ($\mathrm{H}_1 : \pi > 1/6$)
- kritický obor: $Y \geq y_0 = 24$
- padlo nám $Y = 17$, proto (psti binomického rozdělení)

$$p = \mathrm{P}\,(Y \geq 17) = \sum_{k=17}^{100} \binom{100}{k} \left(\frac{1}{6}\right)^k \left(1 - \frac{1}{6}\right)^{100-k} = 0{,}506$$

$$= 1 - \mathrm{P}(Y \leq 16) \qquad \text{[1-BINOMDIST(16;100;1/6;1)]}$$

- protože 50,6 % > 5 %, hypotézu nemůžeme na 5% hladině zamítnout, nemůžeme tvrdit, že pst šestky je větší než 1/6
- neprokázali jsme však, že by hypotéza platila
- na kostce B: $p = \mathrm{P}(Y \geq 41) = 1 - \mathrm{P}(Y \leq 40) = 7{,}4 \cdot 10^{-9}$
  [1-pbinom(40,100,1/6)]

příklad: kostka a oboustranná alternativa

- chceme ověřit, zda je kostka v pořádku
- pokusíme se prokázat, že šestka padla příliš často nebo příliš zřídka (**oboustranná alternativa**)
- $\mathrm{H}_0 : \mathrm{P}(\text{padne šestka}) = 1/6 \qquad (\pi = \pi_0)$
- $\mathrm{H}_1 : \mathrm{P}(\text{padne šestka}) \neq 1/6 \qquad (\pi \neq \pi_0)$
- *proti* hypotéze svědčí malé *nebo* velké hodnoty Y
- pst chyby 1. druhu α rozdělíme na dvě poloviny:
  $\alpha/2$ pro příliš malé Y, $\alpha/2$ příliš velké Y

příklad: kostka, oboustranná alternativa

y_0	8	9	10	...	24	25	26
$P(Y \leq y_0)$	0,010	0,021	0,043	...	0,978	0,988	0,994
$P(Y \geq y_0)$	0,996	0,990	0,979	...	0,038	0,022	0,012
$P(Y = y_0)$	0,006	0,012	0,021	...	0,016	0,010	0,006

- $\alpha/2 = 0{,}025$ ($\alpha/2 = 0{,}05$)
- H_0 zamítneme, když bude $Y \leq 9$ *nebo* když bude $Y \geq 25$
- skutečná pst chyby 1. druhu bude $0{,}021 + 0{,}022 = 0{,}043$
- [pbinom(9,100,1/6)+(1-pbinom(24,100,1/6))] [BINOMDIST(9;100;1/6;1) + 1-BINOMDIST(24;100;1/6;1)]
- hodnoty v rozmezí 10 až 24 (včetně mezí) nesvědčí proti H_0

oboustranná alternativa (přibližně)

- $H_0 : \pi = \pi_0$, např. P(padne šestka) $= 1/6$
- $H_1 : \pi \neq \pi_0$, např. P(padne šestka) $\neq 1/6$
- proti alternativě svědčí Y hodně daleko od $\mu_Y = n\pi_0$ (počítáme za platnosti hypotézy), tj. rel. četnost $f = Y/n$ daleko od π_0
- zavedeme

$$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{f - \pi_0}{\sqrt{\pi_0(1-\pi_0)}}\sqrt{n}$$

- hypotézu zamítneme, bude-li Z daleko od nuly: $|Z| \geq z(\alpha/2)$
- pro $\alpha = 5\ \%$ zamítáme hypotézu, je-li $|Z| \geq 1{,}96$
- $z_A = 0{,}089$ (nezamítneme), $y_B = 6{,}529$ (zamítneme)